

Voting with Their Feet?

Access to Infrastructure and Migration in Nepal

Forhad Shilpi
Prem Sangraula
Yue Li



WORLD BANK GROUP

Development Research Group

Agriculture and Rural Development Team

September 2014

Abstract

Using bilateral migration flow data from the 2010 population census of Nepal, this paper provides evidence on the importance of public infrastructure and services in determining migration flows. The empirical specification, based on a generalized nested logit model, corrects for the non-random selection of migrants. The results show that migrants prefer areas that are nearer to paved roads and have

better access to electricity. Apart from electricity's impact on income and through income on migration, the econometric results indicate that migrants attach substantial amenity value to access to electricity. These findings have important implications for the placement of basic infrastructure projects and the way benefits from these projects are evaluated.

This paper is a product of the Agriculture and Rural Development Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at fshilpi@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Voting with Their Feet? Access to Infrastructure and Migration in Nepal

Forhad Shilpi, Prem Sangraula, Yue Li

JEL classification: O18, R13, R23

Key words: Migration, Public Goods, Income Effect, Amenity Value

Introduction

Do migrants respond to differences in access to public goods and services in addition to income prospects of potential destinations? The income difference between the origin and the destination is the primary factor driving migration in the existing literature on migration (Greenwood (1975) and Borjas (1994), Lall, Selod and Shalizi (2006)). Along with income, recent literature has also highlighted the importance of migration costs as well as migrants' networks in determining migration flows. How provision of public goods and services may influence migration in poorer developing countries remains sparsely studied. This issue is however important in these countries where provision of public goods varies widely across areas. In a Tiebout (1956) sorting model, such disparity in the provision of public goods such as roads, electricity, schools, hospitals, etc. should induce people to "vote with their feet" and to migrate to areas with better access to these infrastructures and services.¹

From a policy perspective, it is important to know how migration responds to the provision of public goods in developing countries for a number of reasons. First, regions within a typical developing country are usually characterized by stark differences in poverty and welfare. Households with poorer attributes such as low levels of education, skills and assets are frequently observed to live in poor areas that are characterized by lack of public infrastructure and services (Shilpi(2011), Dudwick et al. (2011), World Development Report (2009), Kanbur and Venables (2005), Jalan and Ravallion (2002), Ravallion and Wodon (1999)). If migrants do respond to income as well as provision of public infrastructure and services, then migration can act as a powerful instrument in mitigating regional differences in welfare. Second, differential costs of provision of infrastructure and services along with a hard budget constraint often force governments in developing countries to prioritize placement of these public goods. If people do migrate to gain better access to public goods, then the government may be able to rely more on cost considerations to prioritize their placement. Finally, migration in response to public goods and services also has important implications for the way the benefits of public investment are evaluated. A typical evaluation strategy

¹Bayoh, Irwin and Haab (2006) finds that central city's inferior public goods, most notably school quality, play a dominant role in pushing households in the USA metropolitan cities to suburban locations.

of relying on variations in key outcomes such as income or household expenditure across areas with and without public goods would seriously underestimate the benefit of the project. This is because migration in response to a new public good reduces the differences in these outcomes across areas. Using census data from Nepal, this paper provides evidence on the extent to which access to public goods and services influences bilateral migration flow across areas.

The determinants of bilateral migration have been analyzed mostly in the context of international migration (Grogger and Hanson (2011), Ortega and Peri (2013)) and inter-regional migration (Ghatak, Mulhern and Watson (2008), Andrienko and Guriev (2004)).² This literature however focuses primarily on income and migration costs as determinants of migration flow. A recent literature examines how migrants' choice of destination is influenced by locational attributes including the state of public infrastructure and services. For a relatively richer developing country – Brazil – Lall, Timmins and Yue (2009) find that poor migrants are willing to accept lower wages to achieve access to better services while richer migrants are influenced only by income differences.³ Fafchamps and Shilpi (2013) find a statistically significant and numerically large effect of access to paved roads on migrants' destination choice in Nepal: migrants prefer a destination that is closer to a paved road. While contributing to this literature, the analysis in this paper differs from the above papers in a number of ways. Instead of focusing on the destination choice of individual migrants, we analyze bilateral migration flows across multiple sources and destinations. Our empirical specification is derived from a model of utility maximization by the migrants proposed by Ortega and Peri (2013) and Grogger and Hanson (2011). We consider a generalized nested logit model where migrants first decide whether to migrate and then decide among the potential destinations. The advantage of this approach is that the resulting empirical specification includes a correction term for the unobserved heterogeneity between migrants and non-migrants. The above mentioned papers (Fafchamps and Shilpi (2013), Lall, Timmins and Yue (2009)) in contrast side-stepped the issue of migrants' non-random selection

²For a survey of migration literature, please see Greenwood (1975) and Borjas (1994).

³For example, a Brazilian minimum wage worker earning R\$7 an hour was willing to pay R\$420 a year to have access to better health services, R\$87 for a better water supply, and R\$42 for electricity.

by focusing on the choice of destination conditional on migrating. More importantly, we make a distinction between the productivity and amenity values of basic infrastructure and services. For instance, access to electricity allows firms to automate production, shifting the production possibility frontier ("productivity effect"). It also helps households to carry out essential chores efficiently and to enjoy leisure more fully ("amenity effect"). We develop a strategy to uncover the amenity values of infrastructure and services. The strategy relies on a two-stage estimation procedure in which a canonical migration model –ignoring access to infrastructure and services – is fitted at the first stage. The first stage estimation thus allows income to capture the productivity effect of the public goods. To the extent these goods are targeted to more productive areas, income in the first stage estimation picks up that placement effect also. In the second stage, the residual from the first stage is regressed on measures of access to infrastructure and services. By construction, this strategy provides conservative estimates of amenity values of public goods.

The empirical analysis of this paper utilizes the detailed migration information from the population census 2010 of Nepal. Due to the mountainous terrain of the country and limited agricultural potential in many areas, migration is an important livelihood strategy for the Nepalese people. The rough terrain makes the provision of basic infrastructure very difficult with the outcome that large parts of the country are not well served by transport infrastructure. Geographical coverage of electrification remains rather low, serving only a third of rural households. In terms of access to infrastructure and services as well as stage of economic development, Nepal is comparable to many Sub-Saharan African countries. The large geographical variations in access to basic public goods along with vibrant migration flows make Nepal particularly suitable for our study.

When a standard migration model is fitted, our empirical results confirm the common findings of the migration literature that income and distance between source and destination are the two most important determinants of bilateral migration flow. Consistent with the findings of Fafchamps and Shilpi (2013), we find that when measures of access to basic public goods are added as regressors, the magnitude of the income coefficient declines substantially though it still remains statistically significant. This result confirms

that the income coefficient in a standard migration model is likely to be biased upward. Our results show that access to electricity and paved roads are important determinants of migration: migrants prefer areas with better access to electricity and paved roads. The results from the two-stage estimation procedures indicate that migrants attach substantial amenity value to access to electricity as well. Moreover, we find that migrants of different skill levels (primary, secondary and tertiary education level) attach similar amenity values to access to electricity. Thus better access to electricity attracts migrants not only because it brightens their income prospects but also because it offers better quality of life to them.

The rest of the paper is organized as follows. The conceptual framework and empirical specification are presented in Section 2. Section 3 discusses the data. Section 4, organized in subsections, presents the empirical results. Section 5 concludes the paper.

2. Conceptual Framework

2.1 The Model

We start from a simple model of migration where an individual makes a utility maximizing migration decision among multiple destinations within the country. Individual h in her place of residence s decides whether to stay at s or to migrate to any of $i \in I = \{1, \dots, N\}$. Let utility of individual h in location i be denoted U_i^h . Following the literature, we assume that utility U_i^h is a function of the income y_i^h (or consumption) that the individual can achieve in location i , of the prices p_i he or she faces, and a vector of location-specific amenities A_i (Bayoh, Irwin and Haab, 2006). The utility from migrating to a given destination i depends on the migrant's utility from income and amenities suitably adjusted for prices [$u_i^h(y_i^h, A_i, p_i)$] and on the costs C_{si}^h of moving from s to i . Following Grogger and Hanson (2008) and Ortega and Peri (2009), we make a distinction between factors that are shared by all migrants from the same origin and to the same destination, and individual specific factors. The utility in destination i can be expressed as:

$$U_{si}^h = \delta_{si} - \mu_{si}^h = u(y_i, A_i, p_i) - g(C_{si}) - \mu_{si}^h \quad (1)$$

where δ_{si} is an origin-destination specific term shared by all individuals migrating from s to i and μ_{si}^h is the individual migrant specific term. $u_i(y_i, A_i, p_i)$ is the expected utility of individual h in destination i . The expected permanent income of individual h in destination i is the average income y_i . In the empirical estimation, we allow differences in incomes for workers of different skill levels. The expected utility in the destination depends also on the services and amenities available there along with the cost of living. This is important particularly for internal migration where individuals and households may move not only to capture income gain but also to avail themselves of better services and amenities – for instance better schools or health services – at destination. Similarly, C_{si} is the average cost of migration from s to i . The cost term C_{si} captures the physical distance between origin and destination. It also reflects costs incurred by individuals due to social distances (e.g. cultural, ethnic and language differences) between the origin and destination.

We assume that u is an increasing function of y_i , and A_i , and a decreasing function of p_i . We assume that g is an increasing function of C_{si} . Following Grogger and Hanson (2008), we assume that both u and g are approximately linear functions. In the case of no migration, the average expected utility is:

$$U_{ss}^h = \alpha y_s + \beta A_s - \eta p_s$$

where α, β and η are positive constants. The utility from locating in i can be expressed as:

$$U_{si}^h = \alpha y_i + \beta A_i - \eta p_i - \gamma C_{si} - \mu_{si}^h \quad (2)$$

where $\gamma > 0$ is a parameter. The individual specific term μ_{si}^h denotes the idiosyncratic parts of the utility and cost associated with migration by individual h . There is now substantial evidence that migrants may be substantially different from non-migrants in terms of their ability, risk aversion and preferences. Following Ortega and Peri (2009) we assume that:

$$\begin{aligned}
\mu_{si}^h &= (1 - \theta)\varepsilon_{si} \text{ for } i = s \\
&= \lambda^h + (1 - \theta)\varepsilon_{si} \text{ for } i \neq s
\end{aligned} \tag{3}$$

where ε_{si} is iid following an (Weibul) extreme value distribution. λ_i is an individual specific term that affects migrants only and its distribution is assumed to depend on $\theta \in [0, 1]$. Given that ε_{si} has an extreme value distribution, then μ_{si}^h has also an extreme value distribution (Cardell (1991)). The migrant specific term λ_i does not depend on destination and can be thought of capturing the differences in preferences for migration.

Ortega and Peri (2009) show that utility maximization under the distributional assumptions leads to the following condition:

$$\ln S_{si} - \ln S_{ss} - \theta \ln S_{iN} = \alpha[y_i - y_s] + \beta[A_i - A_s] - \eta[p_i - p_s] - \gamma C_{si} \tag{4}$$

where $S_{si} = m_{si}/n_s$, $S_{ss} = (n_s - \sum_{i=1}^N m_{si})/n_s$, and $S_{iN} = m_{si}/\sum_{i=1}^N m_{si}$. n_s is the total native born population of s , m_{si} is the migrants born in s and gone to destination i , and $\sum_{i=1}^N m_{si}$ is the total migrants from s to all possible destinations. S_{si} and S_{iN} are the share of migrants to location i in total native born population of s and total migrants from location s respectively. S_{ss} is the share native born population in s who chose to stay in s . The expression in equation (4) is very similar to an expression under standard logit formulation if $\theta = 0$. The term θ in equation (4) corrects for the differences in utility (due to income, amenity, prices and costs) between migrants and non-migrants (Ortega and Peri (2009)).

Substituting for shares and solving for the logarithm of migration flow ($\ln m_{si}$), equation (4) can be re-written as:

$$\ln m_{si} = \alpha_1 y_i + \beta_1 A_i - \eta_1 p_i - \gamma_1 C_{si} + \tau_s + \xi_{si} \tag{5}$$

where ξ_{si} is the zero-mean measurement error, τ_s is the origin fixed effects and $\alpha_1 = \frac{\alpha}{1-\theta}$, $\beta_1 = \frac{\beta}{1-\theta}$, $\eta_1 = \frac{\eta}{1-\theta}$, and $\gamma_1 = \frac{\gamma}{1-\theta}$. In the standard logit formulation ($\theta = 0$), the fixed effects account for share of the stayers in the population along with income, amenity and prices at the origin [$\tau_s = \ln(n_s - \sum_{i=1}^N m_{si})$]. When migrants differ systematically from non-migrants in preference and ability ($\theta \neq 0$), the fixed effects include a correction term ($\frac{\theta}{1-\theta} \ln \sum_{i=1}^N m_{si}$) for the average unobserved heterogeneity between migrants and non-migrants as well.

We estimate the specification in equation (5) for bilateral gross migration flows among districts in Nepal. Following Grogger and Hanson (2011), we analyze sorting of migrants across destinations. Specifically, we analyze the variations in the skill mix of migrants to different destinations. We define three groups of migrants in terms of their education level: those with (i) less than primary education, (ii) education between primary and secondary levels and (iii) above secondary level.

$$\ln m_{si}^j = \alpha_1^j y_i^j + \beta_1^j A_i - \eta_1^j p_i - \gamma_1^j C_{si} + \tau_s + \xi_{si}^j, \text{ for } j = 1, 2, 3 \quad (6)$$

where j represents the education levels of migrants. The specifications in equations (5) and (6) are based on a linear utility and migration cost functions. A linear formulation can be interpreted as monetary income and cost whereas a log-linear specification would imply as log income and time cost (Ortega and Peri (2013)). We performed estimation using both linear and log linear specifications.

Equations (5) - (6) are the basis of our main empirical estimation. A number of things are worth noting in the estimation of equations (5-6). First, when sufficient numbers of migrants come to a destination, it is expected to have general equilibrium effects on wages, incomes and access to services and amenities.⁴ This would generate a potential endogeneity bias due to the fact that income and amenities in a destination resulted in part from the decision of many migrants to locate there. To eliminate this bias, we use lagged explanatory variables. More precisely, let T be the period for which we have information on explanatory variables and $T + t$ the period at which we observe migrants. Migrants are defined as those who migrated

⁴The effect could be negative – e.g., congestion – or positive – e.g., agglomeration externalities.

between T and $T+t$ whereas explanatory variables come from period T . Limiting the set of migrants in this fashion ensures that migration decisions are based only on information available prior to migration. Second, bilateral migration flows between districts are not always positive. While our main estimation focused on districts with positive migration flows, we also checked the robustness of our results for the sample which included districts with no migration flows. We weight observations by destination population which corrects for potential heteroskedasticity of measurement error. The standard errors are clustered by destination districts to account for within (destination) district correlation of errors.

2.2 Empirical Specification

The basic empirical specifications estimated from the data augment equations (5) and (6) with additional explanatory variables, leading to the following estimating equations:

$$\ln m_{si} = \alpha_1 y_i + \beta_1 A_i - \eta_1 p_i - \gamma_1 C_{si} + \Gamma_1 Z_i + \tau_s + \xi_{si} \quad (7)$$

$$\ln m_{si}^j = \alpha_1^j y_i^j + \beta_1^j A_i - \eta_1^j p_i - \gamma_1^j C_{si} + \Gamma_1 Z_i^j + \tau_s + \xi_{si}^j, \text{ for } j = 1, 2, 3 \quad (8)$$

where Z_i is a vector of locational attributes of destination i . The Z_i vector includes controls for social proximity between source and destination in terms of language, religion and ethnicity. Following standard practice in the literature, we also include a measure of the unemployment rate as a control.

Suppose $\beta_1 = \beta_1^j = 0$, then equations (7 and 8) have the specifications that are comparable to specifications derived from the standard model of determination of migration flows when migrants' preferences for better access to public goods and services are ignored. For simplicity, suppose, y_i and A_i are uncorrelated with rest of the explanatory variables in the above equations and $(\beta_1 \neq 0, \beta_1^j \neq 0)$. We estimate equation (7) ignoring A_i . The estimated coefficient of income in this case is:

$$\bar{\alpha}_1 = \hat{\alpha}_1 + \rho \hat{\beta}_1 \quad (9)$$

where $\bar{\alpha}_1$ is the estimated income coefficient when A_i is ignored and $\hat{\alpha}_1$ and $\hat{\beta}_1$ are the estimated income and amenity/public goods coefficients from the full specification (equation 7) and ρ is the correlation between y_i and A_i . Since income tends to be higher in areas with better public goods, $\rho > 0$. The income coefficient ($\bar{\alpha}_1$) thus overestimates the influence of income differences on migration flow ($\hat{\alpha}_1$) when migrant's preference for public goods is ignored.

The positive correlation between income and public goods means that part of the preference for public goods is due to a preference for higher income. Some of the basic public goods such as roads and electricity have not only direct productivity and hence income effect but also amenity values as they make life easier for households. To explore the amenity value of these goods and services, we utilize a two-stage procedure. At the first stage, we fit a standard migration model ignoring public goods, which allows income to pick up the productivity effect of public goods and services. At the second stage, the residual from the first stage is regressed on the explanatory variables representing access to public goods and services. Similar to equation (9), it follows that:

$$\bar{\beta}_1 = \hat{\beta}_1 - \rho\hat{\alpha}_1$$

where $\bar{\beta}_1$ is the estimate of β_1 from the second stage regression. $\bar{\beta}_1$ thus provides an estimate of the amenity value of public goods and services to migrants.

3. Data

The empirical analysis in this paper draws data from various sources: the population censuses of 2000 and 2010 and the Nepal Living Standard Survey 2002/3. The migration data are collected using the census long form for about 15 percent of the total population. This questionnaire collected information on district of current residence, district of residence 5 years prior to the census, and district of origin. Detailed information is also available on gender, age, education, religion, language spoken, ethnicity and motive for migration. This rich data-set is used to define the gross bilateral migration flows across districts.⁵

⁵Nepal is divided into 75 districts.

The 15 percent sample of population census covers approximately 4.02 million individuals in 740,749 households. Because of our focus on adult migration, we restrict our attention to adults of age 15 and above. Of the 4.04 million individuals, about 35 percent are children below the age of 15 years. Among the adult population (about 2.65 million), about 18.7 percent are living in a district other than their district of birth. Among the migrants, 34 percent have moved in the six years preceding the census, that is, in the period between the 2002/3 NLSS and 2010 census. A large fraction of these individuals have moved for reasons other than work. Marriage is the dominant reason for moving among women (40 percent); study is the dominant reason for moving among children and youths (52 percent). In contrast, of the adult males who migrated during the last 6 years, 54 percent moved for work reasons. We estimate the migration flows among districts using the census data and appropriate population weights. We define two types of migrants: work migrants who moved to seek employment, and all migrants including work migrants as well as those who moved for non-work related reasons. All estimations are carried out for both of these samples.

Figure 1 shows the geographical distribution of migrants in terms of district of residence and origin. As apparent from Figure 1, a small number of destination districts have a high proportion of migrants. In contrast, districts of origin are distributed widely across the country. This reflects the fact that much of the migration is from rural areas to towns and cities. Indeed more than 90 percent of the migrants come from rural areas, and more than half of them migrated to an urban area. The same migration pattern is observed for work migrants.

While the census provided information about migration, it did not collect any information on income, prices and access to services and amenities. We utilize a nationally representative survey of households – the Nepal Living Standard Survey 2002/3 – to derive these explanatory variables. To estimate the average income level in a district i , we ran a regression of the following specification:

$$y_i^h = \delta_i + \zeta x_i^h + v_i^h$$

where y_i^h is the log of income of household h residing in district i , δ_i is the district fixed effects, and x_i^h is a vector of household level explanatory variables and v_i^h is the residual term. The regression includes household size and composition (number of adult males, females and children) as explanatory variables as larger households with more adults tends to earn more income and consume more; omitting them would overestimate incomes in districts where households are larger, e.g., rural districts. Other household characteristics are not included because they could possibly be affected by migration. The estimated δ 's provide of average district income for all households. We also included controls for different education levels to estimate education premia. Incomes for different skill categories are then computed using the education premia.

In the empirical analysis, migration costs are captured by geographical and social distances. For geographical distance between districts, we utilize the arc distance between the district of origin and each possible district of destination, computed from the average longitude and latitude of each district.⁶ We expect the cost and risk of migration to increase with physical distance. Social distance captures the effect of migration networks which are found to be important in determining migration flow (Munshi (1993), Beaman (2012)). Social distance is measured by the index of ethno-linguistic fractionalization (ELF). The ELF index measures the probability that two individuals taken at random belong to the same ethnic or linguistic group. We estimated ELF for each district using the method suggested by Alesina and La Ferrara (2005). The ELF measures are defined for religious, linguistic and ethnic-caste groups using data from the 2000 population census. We computed the district level unemployment rate from the census 2000 data.

Instead of using the share of households with electricity, we construct a measure of electricity connection which does not depend directly on household income. We compute the share of wards – the smallest administrative unit in Nepal – that had electricity connection among all wards in a district using census

⁶The average longitude and latitude of a district are obtained as a weighted average of the longitude and latitude of all the VDC's in the district, where the population of each VDC serves as weight.

2000 data.⁷ This definition of access to electricity avoids the correlation with income that would have resulted from the ability of households with higher incomes to get electricity connection had the access variable been defined at the household level. As a measure of access to markets and other services (schools, hospitals, etc.), we estimated travel time to the nearest paved road from NLSS 2002/2003 data. Travel time to the nearest paved road correlates strongly and positively with other measures of access to services such as travel time to schools, hospitals, local markets and formal banks.

To control for price, we use price of rice. Rice is the most commonly consumed food item in Nepal and thus can be taken as a proxy for the price of common household goods. The NLSS 2002/3 collected information on the quantity and price paid for rice by individual households. We use these data to compute a unit price per kg.

We construct a measure of housing cost using data from the NLSS 2002/3 survey which contained a separate section on housing. The survey collected information on hypothetical and actual house rental values of each household together with house characteristics such as square footage, number and type of rooms, quality of materials, and the availability of various utilities. We use these data to construct a hedonic index of housing costs for each district. Let r_i^k be the house rental price paid (or estimated) by household h in district i and let x_i^h denote a vector of house characteristics. We estimate a regression of the form:

$$\log r_i^k = a_i + bx_i^h + e_i^k$$

where estimate of \hat{a}_i provides a measure the housing cost premium in each district i . In the regression -omitted for the sake of brevity – many of the house characteristics are significant with the expected sign, e.g., larger, better built houses with better in-house amenities are worth more. District price differentials are large and jointly significant. Since the dependent variable is in log form, \hat{a}_i measures the housing cost premium in each district.

Table 1 reports the summary statistics for the dependent and explanatory variables. On average about

⁷There are more than 35 thousand wards in 75 districts in our census long form sample.

45 people migrated between a source-destination pairs. The number of work migrants is smaller – about 19 people. Migration appears to be concentrated at the two ends of the skill distribution: both unskilled (up to primary education) and skilled (above secondary education) tend to migrate at a higher propensity relative to semi-skilled workers (above primary but up to secondary education). This is true for work migration also. This pattern is consistent with the pattern observed for international migration into OECD countries. The propensity to migrate into OECD countries is lower at the semi-skilled level (see Table 1 in Grogger and Hanson (2011)). The median arc distance between source and destination is about 190 km. The average travel time to the nearest paved road is about 7.4 hours indicating relative scarcity of paved roads in Nepal. A large proportion of the country had no electricity connection in 2000, as only a third of the wards in a district reported to have electricity connection. The ELF measures show that Nepal is characterized by low religious diversity but by high ethno-caste diversity. The summary statistics of all other explanatory variables are also reported in Table 1.

4. Empirical Results

The initial set of regression results using the specification in equation (7) are reported in Table 2. All regressions reported in this paper included birth district fixed effects. All regressions are also weighted using destination district population, and all standard errors are clustered at the destination district level to account for any within district error correlations.

4.1. Determinants of Bilateral Migration Flow

We start with the simplest specification which corresponds to the standard specification estimated for bilateral migration flow particularly in the context of international migration (Grogger and Hanson (2011), Ortega and Peri (2013)). The estimation is carried out for two samples: all migrants, and work migrants. The results for this specification are reported in columns 1 and 4 for the all migrant and work migrant samples, respectively. Consistent with the overwhelming evidence from the migration literature, migration flow appears to be associated positively with income at the destination relative to that at source. The income coefficients (columns 1 and 4) are quite precisely estimated. The estimated income coefficient for the

all migrants sample is slightly larger in magnitude than that for the work migrant sample but the hypothesis that the two income coefficients are equal cannot be rejected even at the 20 percent significance level. We introduced up to cubic terms of distance between source and destination, and all three terms are highly statistically significant in explaining variations in migration flow in both samples. The signs of the distance coefficients are consistent with a priori expectation: migration flow declines with an increase in distance between source and destination. Among other explanatory variables, the ELF measures for language and religious diversity have statistically significant coefficients though with opposite signs. The estimated coefficients imply that migration flow to a destination increases with language diversity but decreases with religious diversity and is not significantly associated with ethno-caste diversity. These results are consistent with findings for Nepal reported in Fafchamps and Shilpi (2013). The estimation results show that rice price and unemployment rate do not appear to have statistically significant association with migration flow.

In the next specification, we introduce two measures of public infrastructure and services: travel time to nearest paved road and percentage of wards electrified. The estimated coefficients of these two variables have the expected sign (columns 2 and 5 in Table 2) and are statistically significant at the 1 percent level. The estimated coefficients imply an increase in migration flow to a destination with a decrease in travel time to paved roads and an increase in percentage of wards with electricity connections. Introduction of these two variables led to a substantial decline in the magnitudes of income coefficients: they are now about one third of their respective magnitudes in columns 1 and 4. This confirms our a priori expectation that areas with better access to infrastructure and services are also areas with higher incomes. Despite smaller magnitudes, income is still statistically significant in both the all and work migrants samples.

It is worth noting that the measures of access to public infrastructure and services (travel time to paved road and percentage of wards electrified) are defined from the NLSS 2002/3 and census 2000 data, whereas migration flow is defined over the period 2004 to 2010. Thus for the migration flow under consideration, measures of access to public goods are pre-determined. This helps us to avoid the problem of potential

reverse causation where migration could induce investment in public goods. One remaining concern with the estimated coefficients of access to public goods and services is that they may be picking up the effect of unobserved locational heterogeneity. To redress this issue, we note that in a Roy-Roback model of locational sorting, housing price captures the amenity/dis-amenity value of all location characteristics. As noted by Bayer and Ross (2009), housing price can be taken as a summary measure of the relative attractiveness of an area. We introduce log of housing price premium as a control for unobserved locational heterogeneity where housing price premium estimates come from the NLSS 2002/3. The results from these augmented specifications are reported in columns 3 and 6 for the all migrant and work migrant samples, respectively. The estimated coefficients of housing price premium are positive and statistically significant in both samples suggesting that unobserved locational heterogeneity may be important in determining migration flow. The estimates imply that areas with a higher housing price premium tend to receive higher inflow of migrants in subsequent periods. The estimates of income coefficients have now become somewhat smaller in magnitudes. The same is true for the absolute magnitudes of access to paved roads coefficients. The magnitudes of coefficients of access to electricity on the other hand have increased slightly. More importantly, none of the estimates are statistically significantly different from their respective magnitudes in columns 2 and 5. This suggests that the correlations between unobserved heterogeneity on the one hand and access to paved road and electricity on the other hand are not strong enough to cause any substantial bias in the estimates of coefficients of the latter variables.

Among the other explanatory variables, we find that unemployment rate has the expected negative sign when controls for access to paved road and electricity, and housing price premium are added to the regression. The coefficient of rice price also becomes statistically significant though with a positive sign. Rice price is higher in urban areas compared with rural areas where it is grown because of transportation cost. Rice price thus captures the fact that rural to urban migration is the predominant direction of migration in Nepal. Finally, coefficient estimates are statistically indifferent between the two samples. For the rest of the paper, we thus limit our discussion to results from the full sample. In the next sub-section,

we explore if the results are different for migrants of different skills.

4.2. Determinants of Migration for Different Skill Levels

The determinants of migration flow may be different for people of different skills. To the extent migrants with higher education come from relatively well-off families, they may face lesser credit constraint in financing their migration including the time spell during job search. On the other hand, poorer and unskilled migrants may be pushed out of their source due to adverse shocks and hence their migration may be less sensitive to income differences. To explore these possibilities, we divide migrants into three groups in terms of their education level. Skilled migrants are those with higher than secondary education, and unskilled with primary or less education while semi-skilled belong to the middle group. We report the estimation results for the regression specifications in columns (1) and (3) of Table 2. The regression results are reported in Table 3.

The overall results for all three skill groups are consistent with those reported in Table 2. Some patterns are however worth noting. Income differences between the source and destination seems to have relatively smaller influence on unskilled migrants compared with semi- and skilled migrants, though income coefficients are all positive and statistically significant. The estimates of distance coefficients on other hand display the opposite pattern: they are larger in absolute magnitudes for unskilled and semi skilled migrants compared with skilled migrants. This overall pattern is consistent with the expectations that many more of the unskilled migrants are push migrants and that because of credit constraint, they tend to migrate closer to their origin. Religious diversity – a factor that may relate inversely to migrants’ social network – matters much less for the skilled migrants.

Access to paved roads and electricity have statistically significant coefficients in all three samples. The estimated coefficient is positive for access to electricity and negative for travel time to paved roads. The magnitudes of the coefficients are largest for the skilled migrants who are supposed to be least credit constrained. According to the estimates, unskilled migrants are less sensitive to access to paved roads compared with semi- and skilled migrants: the absolute value of the coefficient of paved roads for unskilled

migrants is about half the size of that for skilled migrants. On other hand, skilled migrants are more sensitive to access to electricity compared with semi-skilled and unskilled migrants. The coefficients of housing price premium have the expected positive sign and statistically significant coefficients in all three regressions. When these locational attributes are added to regressions, the overall pattern in the association between migration flow and income, distance and other variables for the three groups of migrants remain the same. Consistent with our earlier findings, unemployment rate now has statistically significant and negative coefficients in all three regressions. As before, addition of these location characteristics to regressions leads to a significant decline in the magnitudes of income coefficients. While the income coefficients are still statistically significant and have positive signs, their magnitudes are about a third of what they are when access to public goods and housing prices were ignored. This again confirms that income and these locational attributes are significantly and positively correlated. To the extent access to paved roads and electricity contributes to higher income, their respective coefficients capture not only their amenity value but also their productivity effect reflected in higher income. In the following sub-section, we attempt to disentangle their amenity value.

4.3. Migration and Amenity Value of Public Goods and Services

To estimate the amenity value of public goods, we use a two-stage procedure. At the first stage, we estimate a standard migration model ignoring the differences in the provision of electricity or paved roads across areas. This specification thus corresponds to that in column 1 in Table 2, and columns 1, 3 and 5 in Table 3. As shown in equation 9, the coefficient of income in the first stage regression picks up part of the effect of access to public goods and services. At the second stage, the residual from first stage is regressed on the locational attributes. As the first stage regression purges the possible productivity effect of public goods, the second stage estimates thus provide measures of their amenity value. Income in the first stage may pick up more than productivity effect: it may capture part of amenity value that is correlated with productivity effect. Thus second estimates can be considered as lower bound estimates of amenity values.

The second stage estimates are reported in Table 4. The estimates are given for the full sample as well

as for unskilled, semi-skilled and skilled migrants' samples. For each sample, the estimates are reported for two specifications: one excluding housing price premium and the other including it. Regardless of the specifications, the estimates of coefficients of access to electricity fall within a tight interval [0.987-1.215]. These estimates are also statistically significant at 10 percent significance level or less. The estimates of coefficients of travel time to paved road have the expected negative signs but none of them are statistically or numerically significant. The coefficients of housing price premium are also not numerically or statistically significant for any of sub-groups of migrants.

The estimates for access to electricity suggest that migrants do attach amenity value to it. Even after conditioning on income, migration flows are greater to areas which have better access to electricity. The results in Table 4 also suggest no substantial variations in the way migrants of different skill types value access to electricity as an amenity. The estimates of coefficients of access to electricity in Tables 2 and 3 fall within the interval of [2.5-3.12]. The estimates in Table 4 are much smaller in magnitude – roughly about 40 percent of magnitude of estimates in Tables 2 and 3. In other words, of all the migration that happens in response to access to electricity, 40 percent of those is perhaps due to electricity's amenity value.

The estimates in Tables 2 and 3 suggested strong and negative association between bilateral migration flows and travel time to paved road, the estimates in Table 4 show absence of a significant association between these two variables. The strong and negative association between income and geographic isolation (measured here by travel time to paved road) is well noted in the case of Nepal (Fafchamps and Shilpi (2008) and (2013)). The lack of significance of travel time to paved road in the second stage does not necessarily imply that migrants do not attach any amenity value to access to paved road. Rather it suggests that the correlations of travel time to paved road with income and with access to electricity are strong and that given those correlations, it is not possible to disentangle the productivity and amenity value of paved roads.

4.4. Robustness Checks

We perform a number of robustness checks. These checks are conducted for all different samples. To avoid clutter, we report the estimates of the coefficients of access to electricity, paved road and housing price

premium. We also report estimates from two regressions: a full model where all variables are introduced simultaneously; and the estimates from the second stage regression where first stage regression did not include any of the three variables of our interest. The full model thus corresponds to specifications whose results are reported in column 3 of Table 2 and columns 2, 4, and 6 of Table 3. The conditional estimates from second stage correspond to results reported in the even numbered columns of Table 4. We report the results for all migrants in Table 5.

The regression results reported so far come from specifications where income and distance variables are measured in levels. In most migration studies, these variables are often introduced in the logarithmic form. The logarithmic form would imply a log-linear utility function which –according to Grogger and Hanson (2011) – could be mis-specified leading to omitted variable bias. To avoid mis-specification, we estimated the specification consistent with the linear utility function where income and distance enter the regression equation linearly. In the first robustness check, we estimate the regression in log-linear form with both income and distance variables in logarithms. The estimates of parameters of our main interest from the full and conditional (second stage) regressions are reported in columns 1 and 2 of Table 5. The estimates are similar in sign and magnitudes to those reported in Tables 2 and 4.

In the next couple of robustness exercises, we address the issue of potential measurement errors in income estimation. The district level income estimates come from the NLSS 2002/3 data. Income estimates from household surveys typically involve measurement error though NLSS 2002/3 is a nationally representative survey. To check whether our estimates are sensitive to alternative indicators of income, we conduct three robustness checks: (i) Instead of average income adjusted for household size and composition, we use median per capita income at the district level as our income variable. The resulting regression estimates are presented in columns 3 and 4. (ii) While past income from household survey may provide information about potential income at the destination, migrants may not have adequate information on potential income in all different destinations. An important source of information about jobs and incomes for migrants is the past migrants from the same area. We take past stock of migrants who migrated more than 6 years ago

normalized by destination population as an indicator size of migrants' network and add this as a regressor. The results from these regressions are reported in columns 5 and 6. (iii) Incomes across geographical locations are found to be highly correlated with population with income being higher in more populated areas. Population density on the other hand is outcome of migration as well as state of public goods. To avoid reverse causation, we added log of population in 1991 as an additional regressor. The results are reported in columns 7 and 8 in Table 5. The introduction of later two variables (past migrants' stock and population) renders income coefficient smaller in magnitude and statistically significant only at 10 percent level. However, the qualitative results with respect to access to electricity remain unchanged in all of the robustness exercises. When stock of past migrants or population in 1991 are added in the regressions, the magnitudes of coefficients of electricity become somewhat smaller, yet they remain both numerically and statistically significant. The results regarding the significance of access to paved road as a determinant of migration flow become much weaker when past migrants network or past population are added in the regressions. Similar to income, stock of past migrants and population seem to be highly and positively correlated with travel time to paved road. However, the results regarding its amenity value remains unchanged.

In the main regressions, we focused on the sample of all districts with positive migration flows. In the next robustness check, we included all districts including those with zero migration flow. The results are shown in the final two columns of Table 5. The results for this expanded sample is nearly indistinguishable to those reported in Tables 2 and 4.

We repeated the robustness checks for migrants of different skill levels. The results are similar to those reported in Table 5. In the upper panel of Table 6, we report the results when stock of past migrants is included as an additional regressor. The lower panel reports the results with log of 1991 population as an additional regressor. The results are comparable to those for the full sample. Though the magnitudes of coefficients of access to electricity are somewhat smaller compared with those in Table 4, they are all numerically and statistically significant. As before, we find that coefficients of travel time to paved road

become statistically insignificant when population is added as a regressor. In all other cases, migration flow seems to respond significantly with access to paved road in the full regression.

It is worth noting that population of a district captures the relative degree of urbanization as well: districts with larger urban share also have higher population. Since urban areas differ distinctly from rural areas in terms of income and access to public goods, introduction of population in the regression leads to a substantial decline in the magnitudes of coefficients of these variables. Thus inclusion of population as a regressor is likely to bias the estimates of income and public goods coefficients downward. Same argument can be made about stock of past migrants. Our preferred specifications thus exclude these two variables.

4.5. Economic Significance

The explanatory variables in the regressions are measured in different units and thus it is difficult to compare the magnitudes of coefficients of different variables. To provide an idea about the relative importance of different factors in determining migration flows, we provide the estimates of the elasticities based on the estimated coefficients which are reported in Table 7.

We computed elasticities for both the full model which included all explanatory variables simultaneously and the two stage procedure which excluded access to paved road and electricity and housing price premium from the first stage regression. In both models, distance between the source and destination is the most important regressor in terms of the magnitude of its effect on migration. In the full model, other important factors in terms of magnitudes are access to electricity and rice price. Income is also important along with language diversity and access to paved roads, but its magnitude is relatively small implying an increase of migration flow by about 0.4 percent in response to a one percent increase in income. When income is allowed to pick up the effect of infrastructure and services, we find income to be one of the most important determinant of migration flow next only to distance in terms of magnitude of effect. Even after allowing income to pick up part of the effect of access to electricity, access to electricity still remains as an important determinant of migration flow. Interpreting the second stage coefficient as capturing the amenity value, the elasticity of migration with respect to electricity in the second stage confirms that migrants do assign

considerable value to access to electricity as an amenity. Our finding regarding access to electricity is consistent with that of Lall, Timmins and Yue (2009). However, unlike Lall, Timmins and Yue (2009) who find access to electricity to be valued only by the poorer households, our results suggest that its amenity value does not vary across skill groups of migrants. This is perhaps due to the fact that access to electricity is still limited in Nepal with only a third of wards reporting to have access. In contrast, Brazil has nearly universal geographical coverage for electricity (97 percent of rural areas), and the access issue there is more of financial ability to get an electricity connection and paying bills.

5. Conclusions

In the standard new economic geography models, labor is assumed to be mobile in the medium to longer term (Fujita, Krugman and Venables (1999)). With labor mobility, any regionally targeted policy intervention in these models induces labor movement so as to restore spatial equilibrium. Evaluation of large public investment projects such as transportation, electrification and communication on the other hand tends to use spatial variations in outcomes and treatments to estimate returns while ignoring the labor mobility issue. In this paper, we provide evidence on the response of migration to public infrastructure and services using census and household data from a poor developing country, Nepal.

The empirical analysis of this paper incorporates several improvements over the existing literature on the determinants of internal and international migration. The standard model of migration –estimated mostly for international migration – tends to ignore the role of access to public goods and services in the migration decision. Our conceptual model and empirical estimation show that such model tends to over-estimate the importance of income in the determination of migration flow due to the positive correlation between income and provision of public goods. Second, while the empirical studies focusing on migrants’ destination choice do pay attention to spatial differences in the provision of public goods, they tend to side-step the issue of migrants’ non-random selection. There is now a large literature that demonstrates clearly that migrants tend to be different from non-migrants in terms of both observables and unobservables (Gabriel and Schmitz (1995), Dahl (2002), Akee (2006), McKenzie, Gibson and Stillman (2010)). Using

a nested logit model of utility maximization by the migrants – as suggested by Ortega and Peri (2013) – we derive an empirical specification which corrects for the heterogeneity between migrants and non-migrants. Third, we make a distinction between the productivity and hence income effect, and amenity value of basic infrastructure such as electrification. The income effect arises from its direct effect on firm and farm productivity whereas the amenity value derives from its use in household activities (e.g. chores/studying/entertainment). Using the correlation between income and access to these public goods, we develop a strategy to provide conservative estimates of their amenity values.

The empirical results show that migrants prefer areas which are nearer to their birth place and have higher income, better access to electricity and paved roads, higher rice and housing prices and greater language diversity. Consistent with the findings of Fafchamps and Shilpi (2013), we find that when measures of access to basic public goods are added as regressors, the magnitude of income coefficient declines substantially though it still remains statistically significant. This result confirms that the income coefficient in a standard migration model might be biased upward. We find some heterogeneity in the way income, distance and access to a paved road influence migration for different skill groups: more skilled migrants are more responsive to income and access to paved road but less responsive to distance relative to unskilled migrants. The results from the two-stage estimation procedure indicate that migrants attach substantial amenity value to access to electricity. Migrants of different skill levels (primary, secondary and tertiary education level) appear to attach similar amenity values to access to electricity. The results suggest that better access to electricity attracts migrants not only for its positive productivity and income effect but also for its amenity value.

The main finding of this paper that migrants do respond to access to public goods has important implications for the placement and evaluation of basic public infrastructure and services. While geographical coverage of these public goods should be universal, budget constraints often force governments to prioritize their roll out. Our empirical results suggest that governments can perhaps give more weight to cost considerations in prioritizing the roll out. Our results also suggest that impact evaluation of public investment

should pay particular attention to spill-over effects to non-treatment areas due to migration. Such spill-over effects can in turn lead to substantial downward bias in the estimates of returns to public investment when its effect on migration is ignored in the evaluation studies.

References

1. Akee, Randall K.Q., "Deciphering Immigrant Self-Selection: New Evidence from a Developing Country," Kennedy School of Government, Harvard University, Cambridge Mass., 2006 (mimeograph)
2. Andrienko, Y. and S. Guriev, 2004. "Determinants of interregional mobility in Russia: Evidence from panel data", *Economics of Transition*, Volume 12 (1) 2004, 1–27.
3. Bayoh Isaac, Elena G. Irwin and Timothy Haab, "Determinants of Residential Location Choice: How Important Are Local Public Goods in Attracting Homeowners to Central City Locations?," *Journal of Regional Science*, 46(1), 97-120, February 2006.
4. Beaman, Lori A., 2012. "Social Networks and the Dynamics of Labor Market Outcomes: Evidence from Refugees Resettled in the US," *Review of Economic Studies*, vol. 79(1), pages 128-161.
5. Bayer, P. and S. Ross, 2006, "Identifying Individual and Group Effects in the Presence of Sorting: A Neighborhood Effects Application," NBER Working Papers 12211, National Bureau of Economic Research.
6. Borjas, George J., "The Economics of Immigration," *Journal of Economic Literature*, 32(4), 1667-1717, December 1994
7. Dahl, Gordon B., "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets," *Econometrica*, 70(6), 2367-2420, November 2002
8. Dudwick,N., K. Hull, R. Katayama, F. Shilpi and K. Simler, 2011. *From Farm to Firm: Rural-Urban Transition in Developing Countries*, World Bank Direction in Development Series.

9. Fujita, Masahisa, Paul Krugman and Anthony J. Venables, *The Spatial Economy: Cities, Regions, and International Trade*, MIT Press, Cambridge and London, 1999
10. Gabriel, Paul E. and Suzanne Schmitz, "Favorable Self-Selection and the Internal Migration of Young White Males in the United States," *Journal of Human Resources*, 30(3), 460-71, Summer 1995
11. Ghatak, S., A. Mulhern and J. Watson, 2008. "Inter-Regional Migration in Transition Economies: The Case of Poland", *Review of Development Economics*, 12(1), 209–222.
12. Greenwood, Michael J., "Research on Internal Migration in the United States: A Survey," *Journal of Economic Literature*, 13(2), 397-433, June 1975
13. Grogger, Jeffrey and Hanson, Gordon H., 2011. "Income maximization and the selection and sorting of international migrants," *Journal of Development Economics*, vol. 95(1), pages 42-57, May.
14. Harris, John and M. Todaro, "Migration, Unemployment and Development: A Two-Sector Analysis," *Amer. Econ. Rev.*, 60, 126-142, 1970
15. Jalan, Jyotsna and Martin Ravallion, 2002, "Geographic Poverty Traps? A Micro Model of Consumption Growth in Rural China," *Journal of Applied Econometrics*, Vol. 17, p. 329-346.
16. Kanbur, R and A. Venables, 2005, *Spatial Inequality and Development*, Oxford University Press: New York.
17. Lall, Somik and Christopher Timmins, "Rural-Urban Migration: Successful Integration or just "Bright Lights"? Evidence from Brazil and Mexico," Department of Economics, Duke University, 2008 (mimeograph)
18. Lucas, Robert E. Jr, "Life Earnings and Rural-Urban Migration," *Journal of Political Economy*, 112(1), S29-59, Feb. 2004

19. McKenzie, David, John Gibson and Steven Stillman, 2010. "How Important Is Selection? Experimental vs. Non-Experimental Measures of the Income Gains from Migration," *Journal of the European Economic Association*, vol. 8(4), pages 913-945.
20. Munshi, Kaivan, 2003. "Networks in the Modern Economy: Mexican Migrants in the US Labor Market," *Quarterly Journal of Economics*, 118(2), 549-99, May 2003
21. Ortega, F. and G. Peri, 2013. "The Effect of Income and Immigration Policies on International Migration". *Migration Studies*, Volume 1, page 1-28.
22. Ravallion, M and Q. Wodon, 1999, "Poor Areas or Poor People?" *Journal of Regional Sciences*, Vol. 39(4), p. 689-711.
23. Ravallion, Martin and Jyotsna Jalan, 1999, "China's Lagging Poor Areas," *American Economic Review Papers and Proceedings*, Vol. 89(2), pp. 301-305.
24. Roback, J, 1982, "Wages, Rents and Quality of Life," *Journal of Political Economy*, vol. 90, p.1257-78.
25. Roy, A. D, 1951, "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, vol.3, p. 135-46.
26. Shilpi, Forhad. 2011. "Mobility costs and regional inequality: Evidence from Bangladesh," *Journal of Globalization and Development*, 2(1).
27. Tiebout, C. (1956), "A Pure Theory of Local Expenditures", *Journal of Political Economy*, 64 (5): 416-424,
28. World Development Report 2009. *Reshaping Economic Geography*, World Bank, Washington DC.

Figure 1: In and Out- Migration by districts in Nepal, 2011

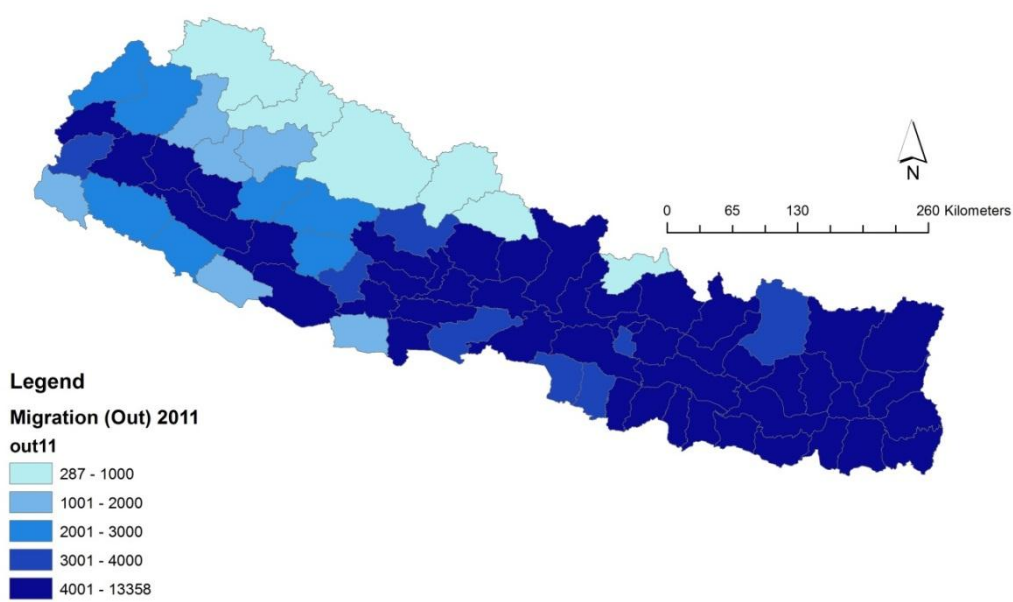
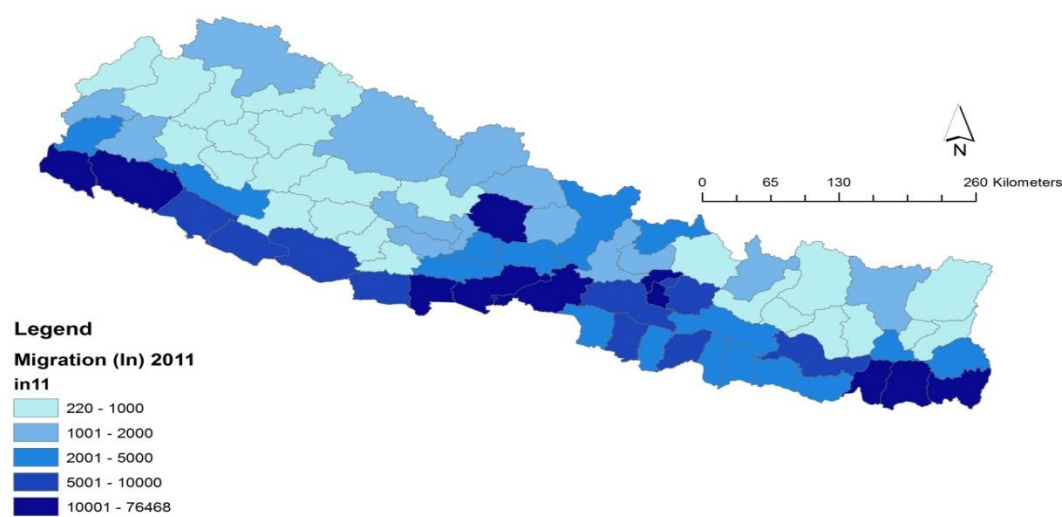


Table 1: Summary Statistics

Variables	Mean	Median	Stand. Dev.	N
Bilateral Flow of				
All Migrants	45.29	5.00	152.81	3434
All Skilled Migrants	14.30	1.00	50.05	3434
All Semi-Skilled Migrants	11.50	1.00	38.99	3434
All Un-Skilled Migrants	19.49	2.00	72.21	3434
Bilateral Flow of				
All Work Migrants	18.64	3.00	58.04	2834
All Skilled Work Migrants	6.89	1.00	24.36	2834
All Semi-Skilled Work Migrants	4.57	0.99	15.80	2834
All Un-Skilled Work Migrants	7.18	2.00	21.12	2834
Monthly Household Income (Rs 000)	1.33	1.29	0.63	3434
Price of Rice (Rs per kg)	21.43	19.46	6.20	3434
Unemployment Rate	1.62	1.57	0.79	3434
Ethno-Language Fractionalization Index: Language	0.45	0.48	0.22	3434
Ethno-Language Fractionalization Index: Religion	0.28	0.28	0.17	3434
Ethno-Language Fractionalization Index: Caste	0.82	0.83	0.11	3434
Distance between source and destination (km)	236	190	168	3434
Proportion of wards with Electricity	0.35	0.33	0.22	3434
Travel time to nearest paved road (hour)	7.43	1.66	12.68	3434
Log(housing price premium)	1.63	1.78	0.80	3434

Table 2: Determinants of Migration Flow: Regression Results

	(1)	(2)	(3)	(4)	(5)	(6)
	Log(All Migrants)			Log(All Work Migrants)		
Income	1.234*** (3.796)	0.383** (2.200)	0.314* (1.954)	1.021*** (3.752)	0.323** (2.133)	0.281** (1.995)
Price of Rice	0.0336 (1.283)	0.110*** (2.949)	0.0954** (2.621)	0.0420* (1.844)	0.111*** (3.232)	0.100*** (2.984)
Unemployment Rate	0.157 (0.891)	-0.158 (-1.410)	-0.219** (-2.231)	0.0806 (0.534)	-0.208* (-1.865)	-0.258** (-2.560)
ELF-Language	3.144*** (7.350)	1.802*** (3.070)	1.546*** (2.792)	2.727*** (6.741)	1.605*** (3.143)	1.411*** (2.847)
ELF-Religion	-2.205*** (-2.764)	-2.098** (-2.285)	-1.740** (-2.003)	-2.118*** (-3.074)	-2.145** (-2.642)	-1.915** (-2.469)
ELF-Caste/Ethnicity	-0.451 (-0.310)	0.344 (0.175)	0.632 (0.333)	-0.0888 (-0.0638)	0.774 (0.404)	1.040 (0.557)
Proportion of wards with Electricity		2.853*** (4.191)	3.125*** (4.896)		2.314*** (3.734)	2.515*** (4.265)
Travel time to nearest paved road		-0.0397*** (-2.728)	-0.0370*** (-2.651)		-0.0352*** (-2.708)	-0.0336** (-2.628)
Log(housing price premium)			0.273*** (2.688)			0.191* (1.952)
Distance	-0.0313*** (-16.29)	-0.0319*** (-17.38)	-0.0317*** (-17.16)	-0.0259*** (-16.54)	-0.0267*** (-18.25)	-0.0268*** (-18.14)
Distance Squared/1000	0.0699*** (11.93)	0.0713*** (12.70)	0.0707*** (12.42)	0.0579*** (11.81)	0.0599*** (12.32)	0.0602*** (12.28)
Distance Cubed/1000000	-0.0507*** (-10.11)	-0.0521*** (-11.45)	-0.0517*** (-10.96)	-0.0421*** (-9.760)	-0.0437*** (-10.12)	-0.0441*** (-10.16)
Intercept	2.724 (1.555)	1.951 (0.941)	1.648 (0.832)	1.792 (1.161)	0.969 (0.491)	0.718 (0.376)
Observations	3,434	3,434	3,434	2,834	2,834	2,834
R-squared	0.658	0.733	0.741	0.617	0.685	0.690

Note: All regressions include birth district fixed effects and weighted using destination population. All standard errors are clustered at destination district level. ELF: Ethno-Linguistic Fractionalization Index.

Robust t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3: Determinants of Migration Flow by Skill Groups: Regression Results

	Log(All Migrants)					
	Unskilled		Semi-Skilled		Skilled	
	(1)	(2)	(3)	(4)	(5)	(6)
Income	0.963*** (3.559)	0.246** (2.220)	1.017*** (3.719)	0.267** (2.127)	1.236*** (3.635)	0.328* (1.973)
Price of Rice	0.0301 (1.374)	0.0653** (2.194)	0.0330 (1.499)	0.0872*** (2.794)	0.0447* (1.709)	0.113*** (2.981)
Unemployment Rate	0.146 (1.160)	-0.132* (-1.852)	0.0956 (0.688)	-0.193** (-2.331)	0.175 (0.974)	-0.185* (-1.734)
ELF-Language	2.541*** (7.304)	1.304*** (2.947)	2.463*** (6.592)	1.230*** (2.769)	2.779*** (6.272)	1.268** (2.372)
ELF-Religion	-2.284*** (-3.708)	-1.912*** (-2.787)	-2.031*** (-3.148)	-1.722** (-2.480)	-1.700** (-2.108)	-1.270 (-1.446)
ELF-Caste/Ethnicity	0.296 (0.240)	0.998 (0.576)	0.499 (0.402)	1.260 (0.745)	-0.762 (-0.534)	0.307 (0.168)
Proportion of wards with Electricity		2.558*** (4.857)		2.504*** (4.702)		2.925*** (4.656)
Travel time to nearest paved road		-0.0210** (-2.008)		-0.0308*** (-2.747)		-0.0401*** (-2.754)
Log(housing price premium)		0.169** (2.301)		0.156** (2.039)		0.267** (2.616)
Distance	-0.0327*** (-17.13)	-0.0332*** (-17.82)	-0.0298*** (-17.70)	-0.0302*** (-18.62)	-0.0275*** (-15.32)	-0.0278*** (-16.34)
Distance Squared/1000	0.0751*** (12.27)	0.0762*** (12.41)	0.0695*** (12.59)	0.0703*** (12.85)	0.0648*** (11.71)	0.0653*** (12.23)
Distance Cubed/1000000	-0.0542*** (-9.808)	-0.0553*** (-10.11)	-0.0506*** (-10.12)	-0.0515*** (-10.59)	-0.0484*** (-10.23)	-0.0491*** (-11.03)
Intercept	2.172 (1.490)	1.656 (0.922)	1.191 (0.824)	0.412 (0.236)	1.344 (0.781)	0.150 (0.0773)
Observations	3,434	3,434	3,434	3,434	3,434	3,434
R-squared	0.627	0.690	0.611	0.685	0.630	0.716

Note: All regressions include birth district fixed effects and weighted using destination population. All standard errors are clustered at destination district level. . ELF: Ethno-Linguistic Fractionalization Index. Robust t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Amenity Value of Public Infrastructure and Services

	All		Unskilled		Semi-Skilled		Skilled	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Proportion of wards with Electricity	1.215**	1.184**	1.018**	1.013**	0.989**	0.987**	1.111*	1.077*
	(2.293)	(2.364)	(2.215)	(2.269)	(2.177)	(2.233)	(1.888)	(1.967)
Travel time to nearest paved road	-0.00622	-0.00705	-0.00301	-0.00315	-0.00500	-0.00506	-0.00692	-0.00782
	(-0.472)	(-0.551)	(-0.309)	(-0.330)	(-0.468)	(-0.483)	(-0.542)	(-0.637)
Log(housing price premium)		0.0715		0.0119		0.00580		0.0778
		(0.520)		(0.119)		(0.0552)		(0.572)

Note: All standard errors are clustered at destination district level.

Robust t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 5: Robustness Checks

	Log-linear Specification		Median Per Capita Income		Log(Migrants) Migrants' Stock at destination		Ln(population91)		Expanded sample	
	Full	2nd stage	Full	2nd stage	Full	2nd stage	Full	2nd stage	Full	2nd stage
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Log-linear Specification										
Proportion of wards with Electricity	3.308*** (5.169)	1.590** (2.477)	3.140*** (5.251)	1.325*** (3.063)	2.568*** (4.561)	0.919** (2.171)	2.713*** (4.308)	0.736** (2.368)	3.084*** (5.153)	1.334** (2.413)
Travel time to nearest paved road	-0.0405** (-2.562)	-0.0103 (-1.181)	0.0319** (-2.247)	-0.00437 (-0.462)	-0.0338*** (-2.746)	-0.00673 (-0.651)	-0.0132 (-0.928)	0.000572 (0.0585)	0.0356*** (-2.815)	-0.00563 (-0.453)
Log(housing price premium)	0.304*** (2.991)	0.0815 (0.548)	0.272** (2.592)	0.0711 (0.463)	0.240*** (2.676)	0.0732 (0.638)	0.325*** (3.193)	0.186 (1.650)	0.207** (2.589)	0.0565 (0.529)

Note: All regressions include birth district fixed effects and weighted using destination population. All standard errors are clustered at destination district level.

Robust t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 6: Robustness Checks (Skill Levels)

	Unskilled		Semi-Skilled		Skilled	
	Full	2nd stage	Full	2nd stage	Full	2nd stage
	(1)	(2)	(3)	(4)	(5)	(6)
Migrants' Stock at destination						
Proportion of wards with Electricity	1.909*** (4.377)	0.721** (1.995)	1.883*** (4.296)	0.712* (1.964)	2.365*** (4.303)	0.821* (1.717)
Travel time to nearest paved road	-0.0171** (-2.001)	-0.00269 (-0.390)	-0.0272*** (-2.954)	-0.00474 (-0.600)	-0.0369*** (-2.817)	-0.00757 (-0.764)
Log(housing price premium)	0.131** (2.248)	0.0133 (0.179)	0.120* (1.913)	0.00710 (0.0878)	0.234** (2.588)	0.0794 (0.695)
Ln(population91)						
Proportion of wards with Electricity	2.238*** (4.369)	0.661** (2.413)	2.172*** (4.268)	0.622** (2.285)	2.486*** (4.058)	0.644** (2.028)
Travel time to nearest paved road	-0.00259 (-0.263)	0.00284 (0.389)	-0.0116 (-1.088)	0.00123 (0.159)	-0.0145 (-1.057)	- (-0.0305)
Log(housing price premium)	0.209*** (2.764)	0.0989 (1.237)	0.198** (2.535)	0.0992 (1.169)	0.323*** (3.133)	0.192* (1.735)

Note: All regressions include birth district fixed effects and weighted using destination population. All standard errors are clustered at destination district level.

Robust t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 7: Relative Magnitudes of Estimated Coefficients

	Coefficient Estimates			Elasticity		
	Full Model	First Stage	Second Stage	Full Model	First Stage	Second Stage
Income	0.314*	1.234***		0.418	1.642	
Price of Rice	0.0954**	0.0336		2.044	0.720	
Unemployment Rate	-0.219**	0.157		-0.356	0.255	
ELF-Language	1.546***	3.144***		0.689	1.402	
ELF-Religion	-1.740**	-2.205***		-0.490	-0.621	
ELF-Caste/Ethnicity	0.632	-0.451		0.516	-0.368	
Proportion of wards with Electricity	3.125***		1.184**	1.097		0.416
Travel time to nearest paved road	-0.0370***		-0.00705	-0.275		-0.052
Log(housing price premium)	0.273***		0.0715	0.273		0.072
Distance	-0.0317***	-0.0313***		-7.478	-7.384	
Distance Squared/1000	0.0707***	0.0699***		5.925	5.858	
Distance Cubed/1000000	-0.0517***	-0.0507***		-1.949	-1.912	
Total Distance Effect				-3.503	-3.438	

Note: All regressions include birth district fixed effects and weighted using destination population. All standard errors are clustered at destination district level. . ELF: Ethno-Linguistic Fractionalization Index. Robust t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1